Statistical Methods in Particle Physics

3. Uncertainties

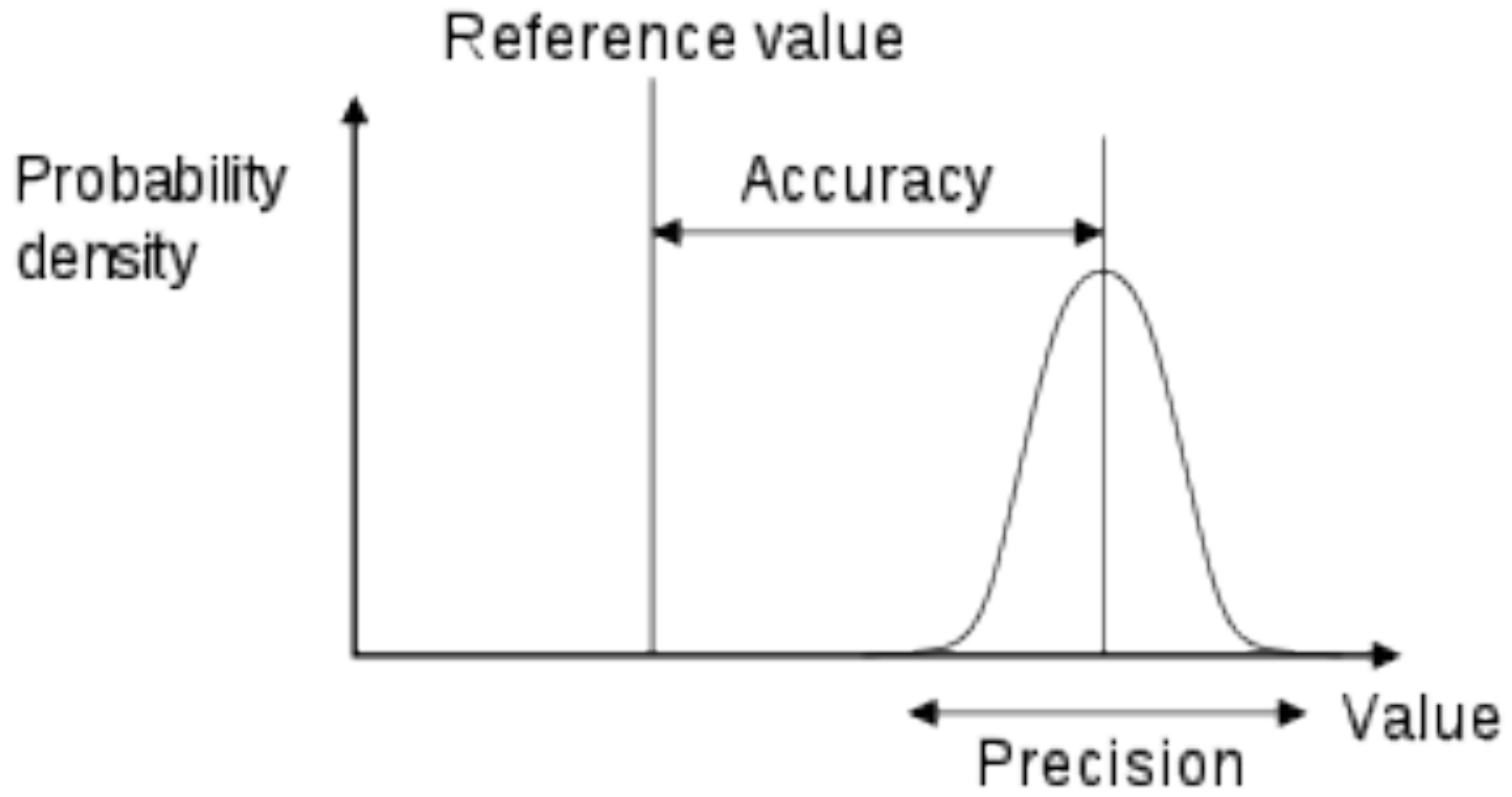
Prof. Dr. Klaus Reygers (lectures)
Dr. Sebastian Neubert (tutorials)

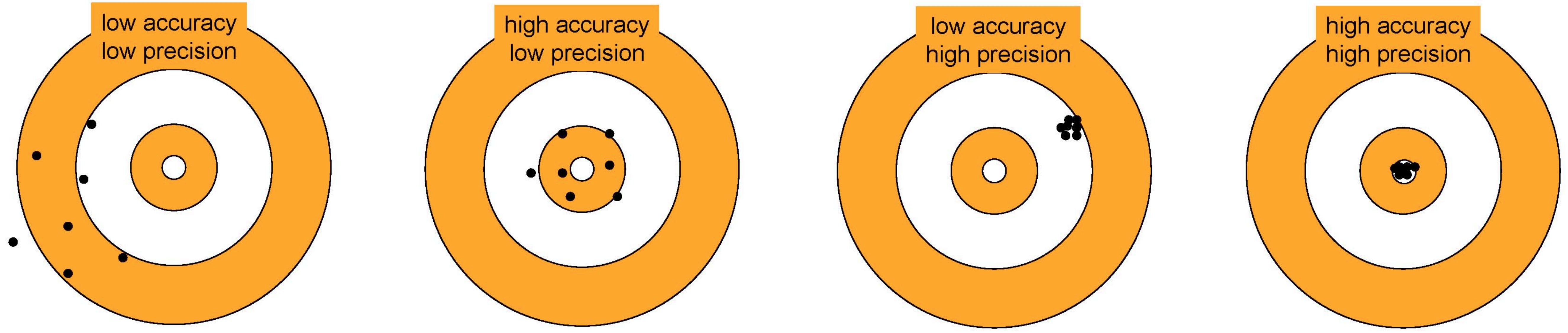
Heidelberg University

WS 2017/18

Statistical and Systematic Uncertainties

Precision and Accuracy





Ways to Quote Uncertainties

$$t = (34.5 \pm 0.7)\ 10^{-3}\ \mathrm{s}$$
$$t = 34.5\ 10^{-3}\ \mathrm{s} \pm 2\,\%$$
$$x = 10.3^{+0.7}_{-0.3}$$
$$m_e = (0.510\,999\,06 \pm 0.000\,000\,15)\ \mathrm{MeV}/c^2$$
$$m_e = 0.510\,999\,06\ (15)\ \mathrm{MeV}/c^2$$
$$m_e = 9.109\,389\,7\,10^{-31}\,\mathrm{kg}\ \pm 0.3\,ppm$$

An uncertainty σ represents some kind of probability distribution (often a Gaussian, if not stated otherwise)

If no further information is given the interval x ± σ corresponds to a a probability of 68% ("1σ errors")

Statistical and Systematic Uncertainties

$$x = 2.34 \pm 0.05\,(\mathrm{stat.}) \pm 0.03\,(\mathrm{syst.})$$

quoting stat. and syst. uncertainty separately gives us an idea whether taking more data would be helpful

Statistical or random uncertainties

- Uncertainties that can be reliably estimated by repeating measurements
- They follow a known distribution like a Poisson rate or are determined empirically from the distribution of an unbiased, sufficiently large sample.
- Relative uncertainty reduces as $1/\sqrt{N}$ where N is the sample size

Systematic uncertainties

- Cannot be calculated solely from sampling fluctuations
- In most cases don't reduce as $1/\sqrt{N}$ (but often also become smaller with larger N)
- Difficult to determine, in general less well known than the statistical uncertainty
- Systematic uncertainties $\neq$ mistakes
(a bug in your computer code is not a systematic uncertainty)

Statistical Uncertainties: Examples

Radioactive decays (→ Poisson distribution)

- You measure N = 150 decays.
- The result is reports as $N \pm \sqrt{N} \approx 150 \pm 12$

Efficiency of a detector (→ Binomial distribution)

- From N_0 = 60 particles which traverse a detector, 45 are measured
- $\varepsilon = N/N_0 = 0.75$

$$\sigma_N^2 = N_0 \varepsilon (1-\varepsilon) \quad \rightsquigarrow \quad \sigma_\varepsilon = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N_0}} = \sqrt{\frac{0.75 \cdot 0.25}{60}} = 0.06$$

Systematic Uncertainties: Examples

- Calibration uncertainties of the measurement apparatus
  - E.g., energy scale uncertainty of a calorimeter
- Uncertainty of the detector resolution
- Detector acceptance
- Limited knowledge about background processes
- Uncertainties of auxiliary quantities
  - E.g. reference branching ratios uses as input
  - Uncertainty of theoretical quantities
- …

A large fraction of the work in a particle physics analyses is estimating systematic uncertainties!

How to Deal with Systematic Uncertainties?

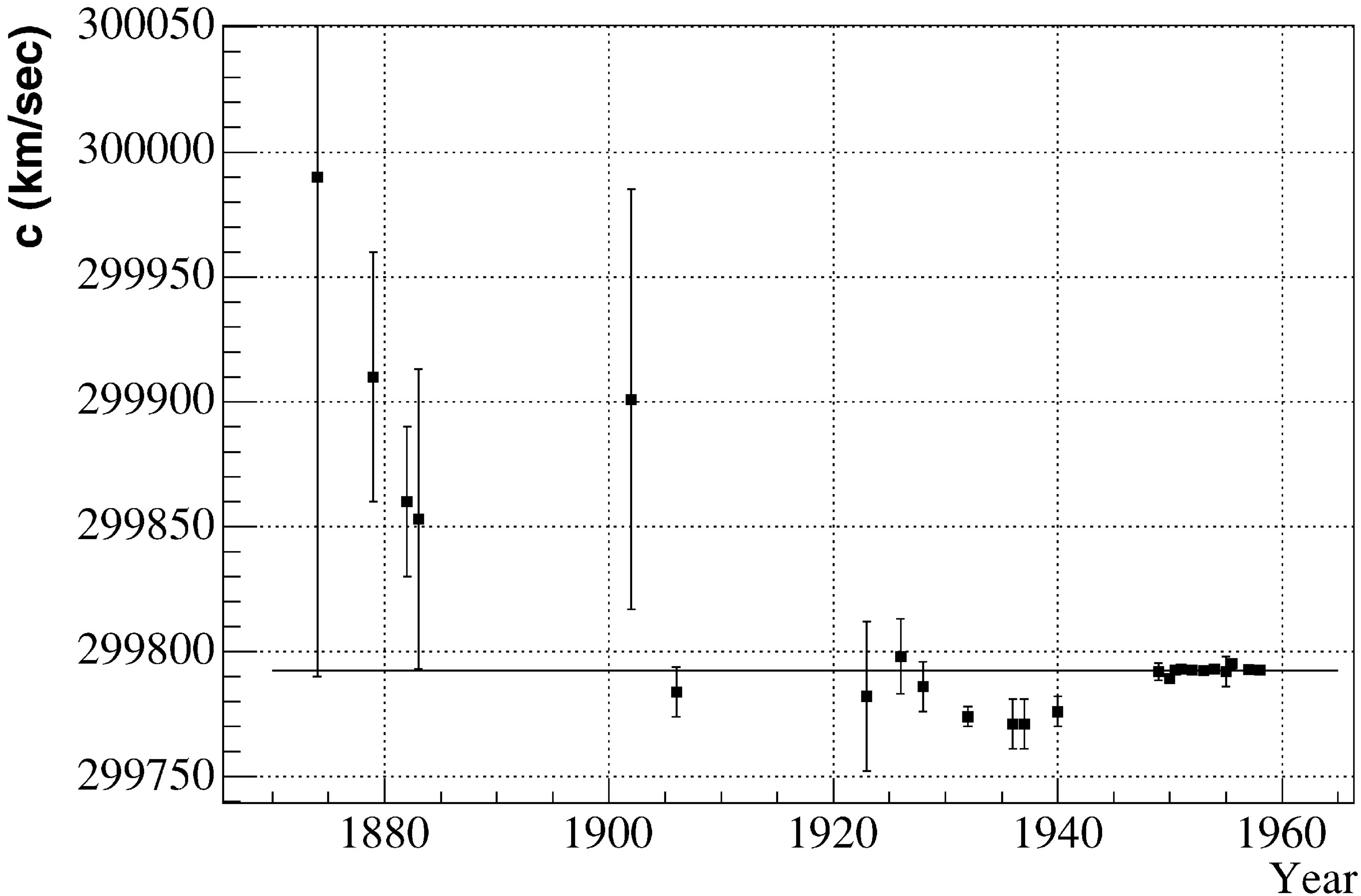
Top-Down Approach

- Think about all possible sources of potential systematics
- Requires experience

Bottom-Up Approach

- Try to find systematic uncertainties not considered in top-down approach
- Internal cross checks
- Split data into independent subsets
- Compare independent analyses if possible
- Cut variation:
  - helps to identify systematics uncertainties
  - but reasons for possible differences should be understood
  - often difficult to separate statistical fluctuations from real systematic effects

Speed of Light vs. Year of Publication

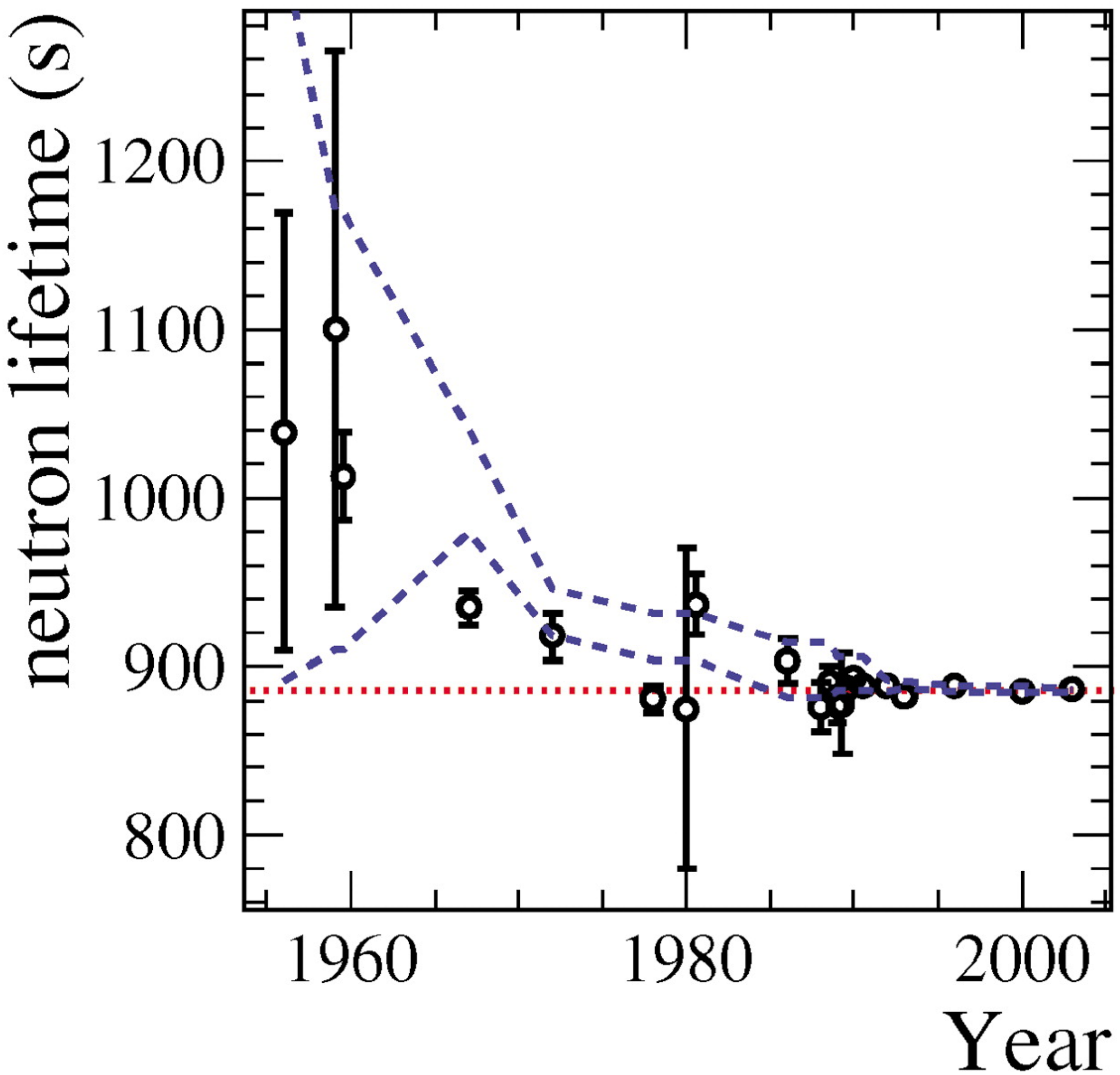


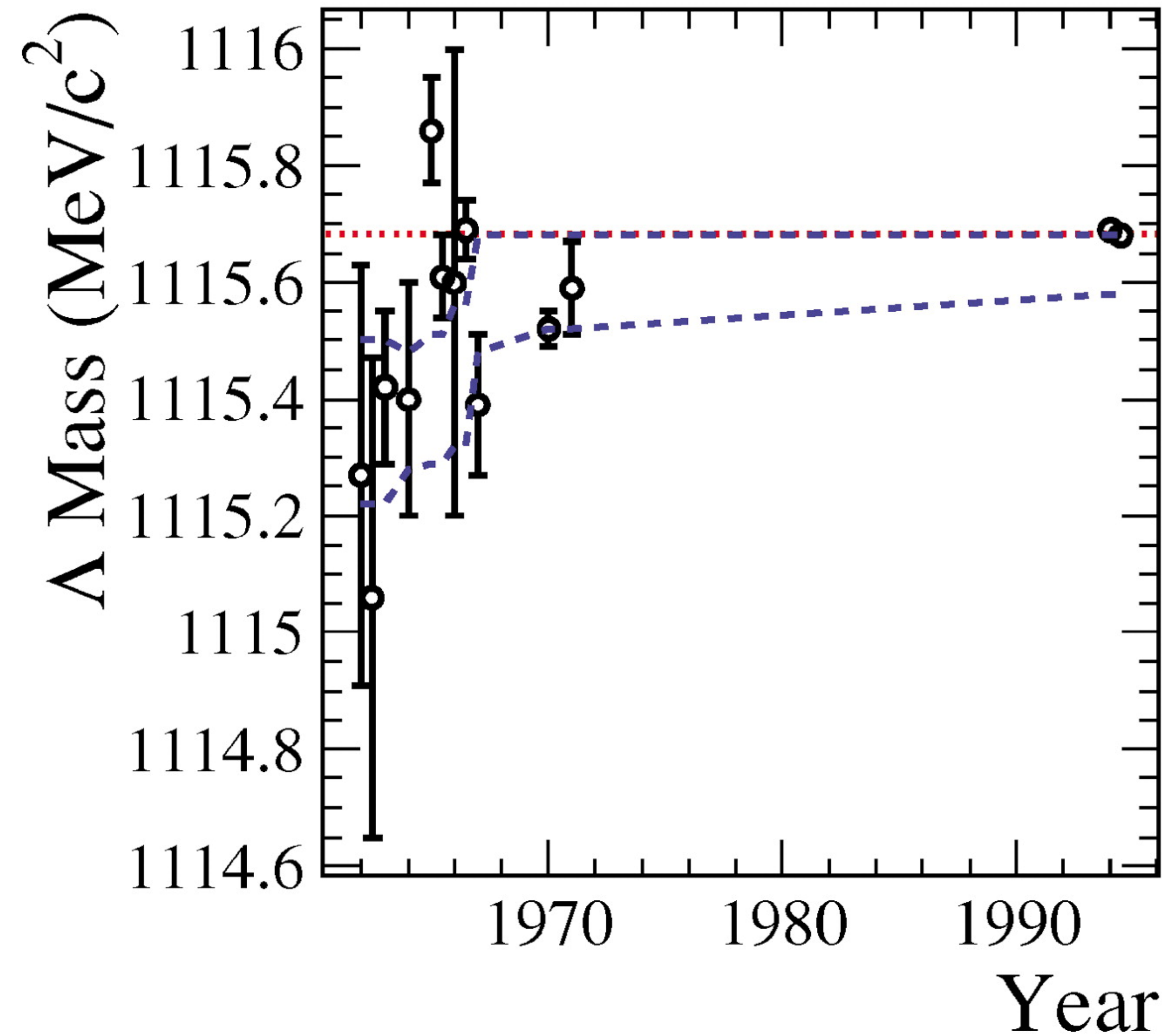
Klein JR, Roodman A. 2005.
Annu. Rev. Nucl. Part. Sci. 55:141–63

Experimenter's Bias?

Klein JR, Roodman, A. 2005,
Annu. Rev. Nucl. Part. Sci. 55:141–63

Do researches unconsciously work toward a certain value?





Possible bias:

the investigator searches for the source or sources of such errors, and continues to search until he gets a result close to the accepted value.

Then he/she stops!

Blind Analyses

Klein JR, Roodman, A. 2005,
Annu. Rev. Nucl. Part. Sci. 55:141–63

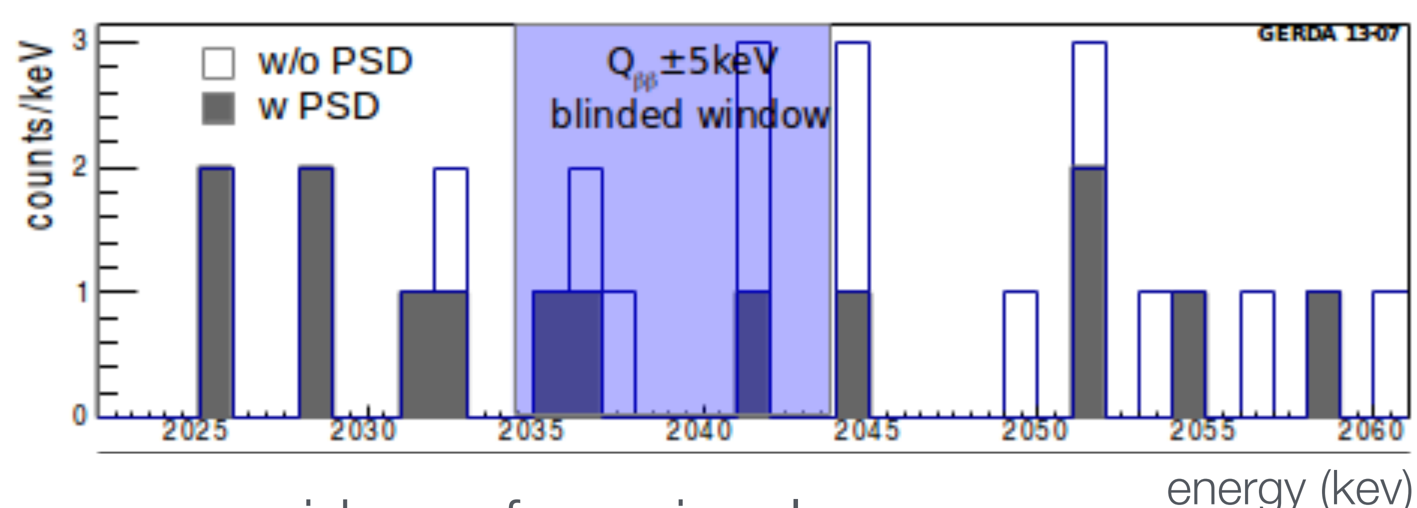
Avoid experimenter's bias by hiding certain aspects of the data.

Things that can be hidden in the analysis:

- The signal events, when the signal occurs in a well-defined region of the experiment's phase space.
- The result, when the numerical answer can be separated from all other aspects of the analysis.
- The number of events in the data set, when the answer relies directly upon their count.
- A fraction of the entire data set.

Example: GERDA experiment

- search for neutrinoless double beta decay
- Signal: sharp peak
- Background model fixed prior to unblinding of signal region



→ no evidence for a signal

Combination of Systematic Uncertainties

In most cases one tries to find independent sources of systematic uncertainties. These independent uncertainties are therefore added in quadrature:

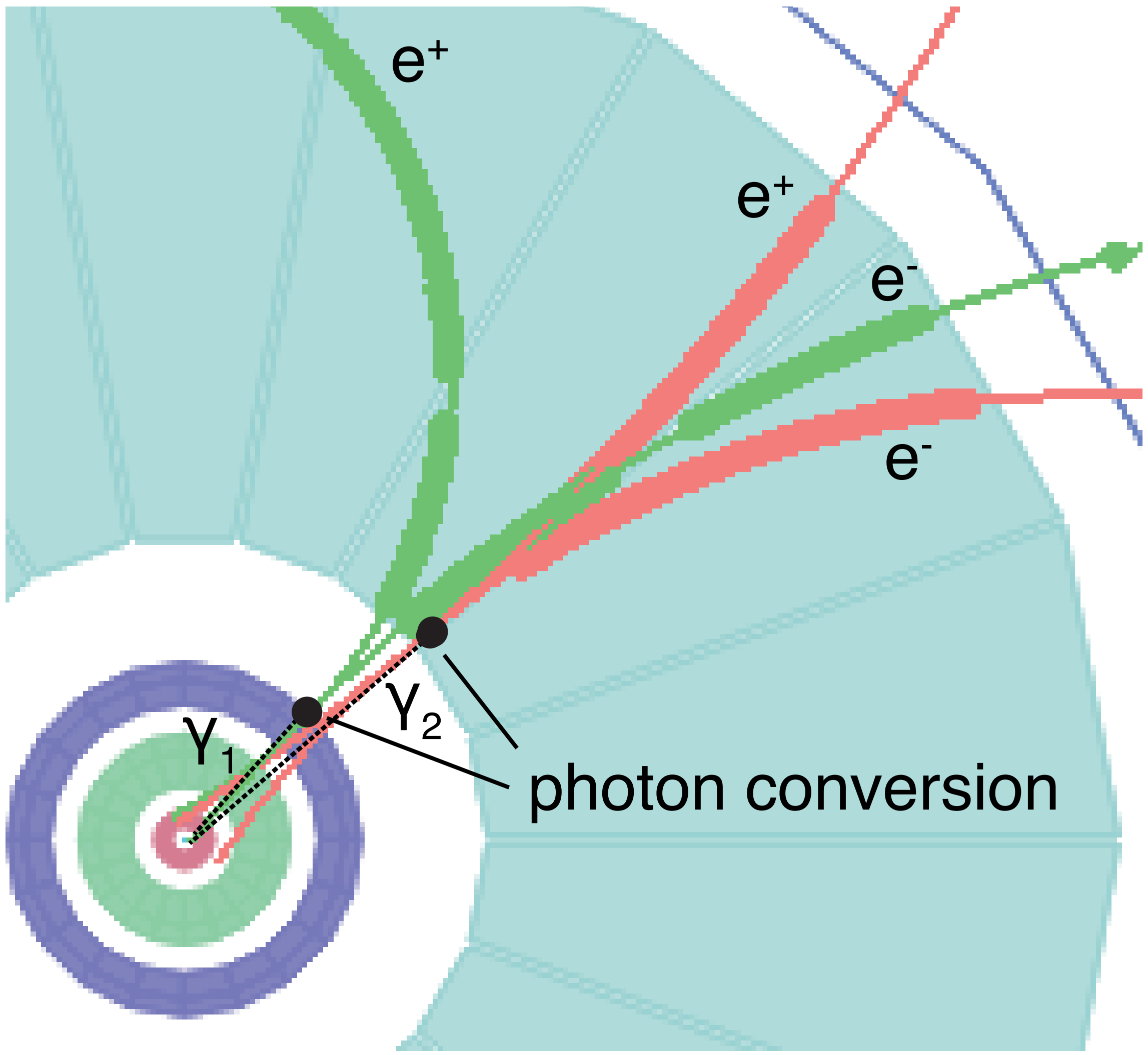
$$\sigma_{\mathrm{tot}}^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$$

Often a few source dominate the systematic uncertainty

→ No need to work to hard on correctly estimating the small uncertainties

Example:
Neutral Pions Yields from Converted Photons in ALICE

$$\pi^0 \rightarrow \gamma + \gamma, \qquad \gamma + \text{material} \rightarrow e^+ + e^-$$



	PCM	
	pp	
	1.1 GeV/c	5.0 GeV/c
Material budget	9.0	9.0
Yield extraction	0.6	2.6
e^+/e^- identification	0.7	1.4
Photon identification ($\chi^2(\gamma)$)	2.4	0.9
π^0 reconstruction efficiency	0.5	3.6
Pile-up correction	1.8	1.8
Total	9.5	10.3

In this measurement the material budget uncertainty dominates the systematic uncertainty

Describing Correlated Systematic Uncertainties (I)

Consider two measurement x_1 and x_2 with with individual random uncertainties $\sigma_{1,r}$ and $\sigma_{2,r}$ and a common systematic uncertainty σ_s:

$$x_i = x_\mathrm{true} + \Delta x_{i,\mathrm{r}} + \Delta x_\mathrm{s}$$

$$\langle \Delta x_{i,\mathrm{r}} \rangle = 0, \quad \langle \Delta x_\mathrm{s} \rangle = 0,$$
$$\langle (\Delta x_{i,\mathrm{r}})^2 \rangle = \sigma_{i,\mathrm{r}}^2, \quad \langle (\Delta x_\mathrm{s})^2 \rangle = \sigma_s^2$$

Variance:

$$\begin{aligned} V[x_i] &= \langle x_i^2 \rangle - \langle x_i \rangle^2 \\ &= \langle (x_\mathrm{true} + \Delta x_{i,\mathrm{r}} + \Delta x_\mathrm{s})^2 \rangle - \langle x_\mathrm{true} + \Delta x_{i,\mathrm{r}} + \Delta x_\mathrm{s} \rangle^2 \\ &= \langle (\Delta x_{i,\mathrm{r}} + \Delta x_\mathrm{s})^2 \rangle \\ &= \sigma_{i,r}^2 + \sigma_s^2 \end{aligned}$$

Covariance:

$$\begin{aligned} \mathrm{cov}[x_1, x_2] &= \langle x_1 x_2 \rangle - \langle x_1 \rangle \langle x_2 \rangle \\ &= \ldots \\ &= \sigma_s^2 \end{aligned}$$

Describing Correlated Systematic Uncertainties (II)

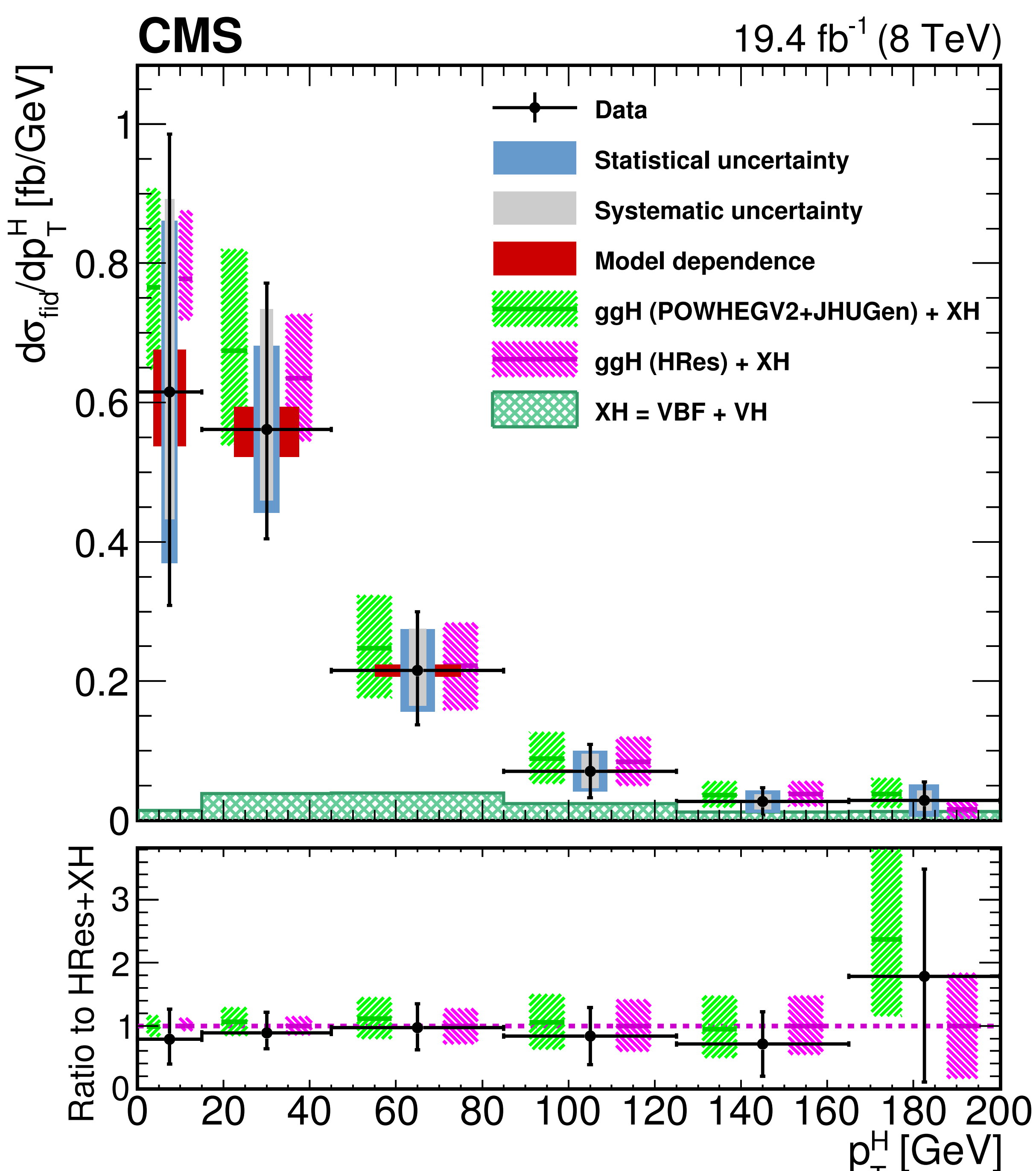
Covariance matrix for x_1 and x_2:

$$V = \begin{pmatrix} \sigma_{1,r}^2 + \sigma_s^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_{2,r}^2 + \sigma_s^2 \end{pmatrix}$$

This also works when the uncertainties are quoted as relative uncertainties:

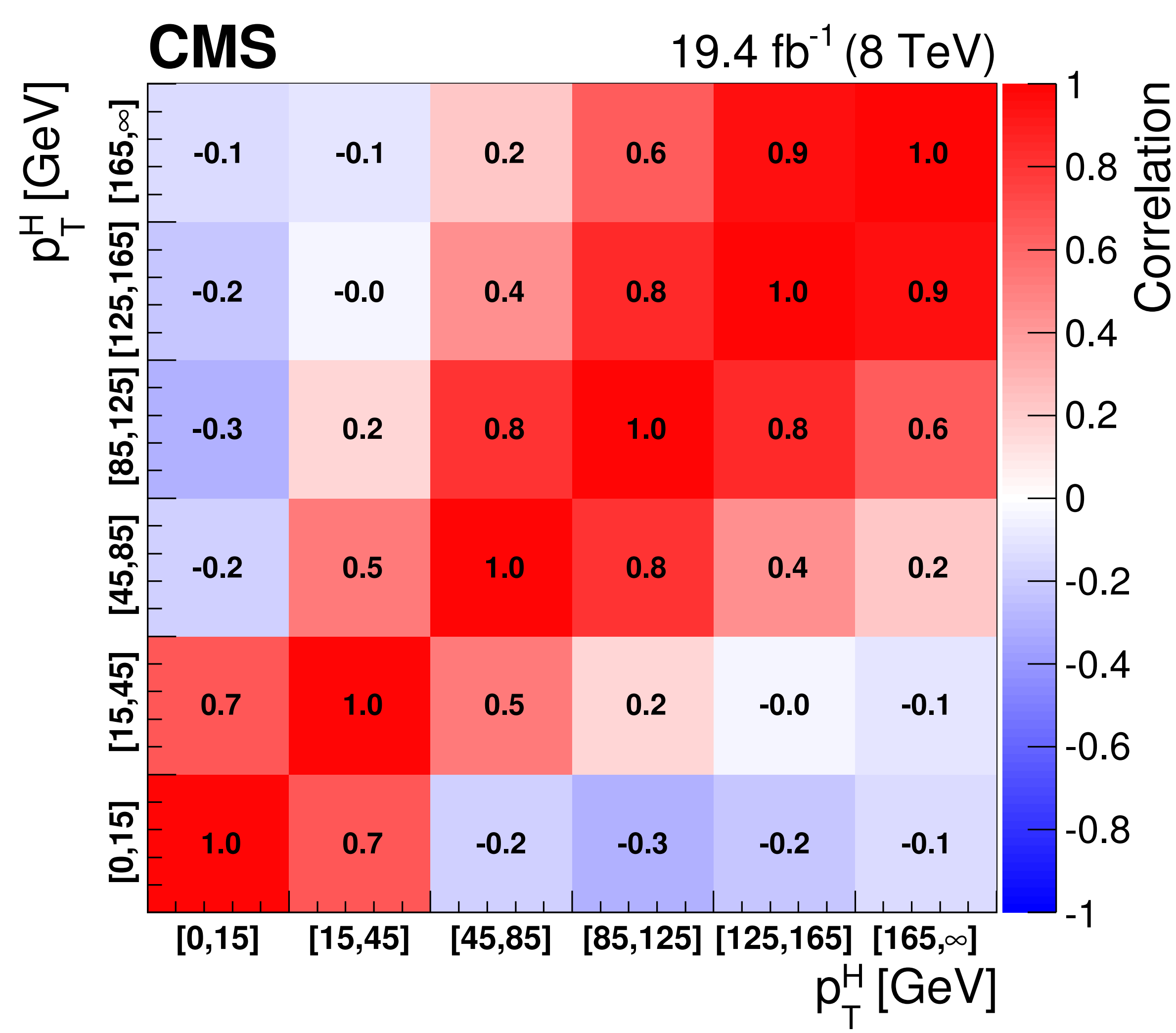
$$\sigma_s = \varepsilon x \qquad \rightsquigarrow \qquad V = \begin{pmatrix} \sigma_{1,r}^2 + \varepsilon^2 x_1^2 & \varepsilon^2 x_1 x_2 \\ \varepsilon^2 x_1 x_2 & \sigma_{2,r}^2 + \varepsilon^2 x_1^2 \end{pmatrix}$$

Example:
Transverse Momentum Spectrum of the Higgs-Boson



arXiv:1606.01522v1

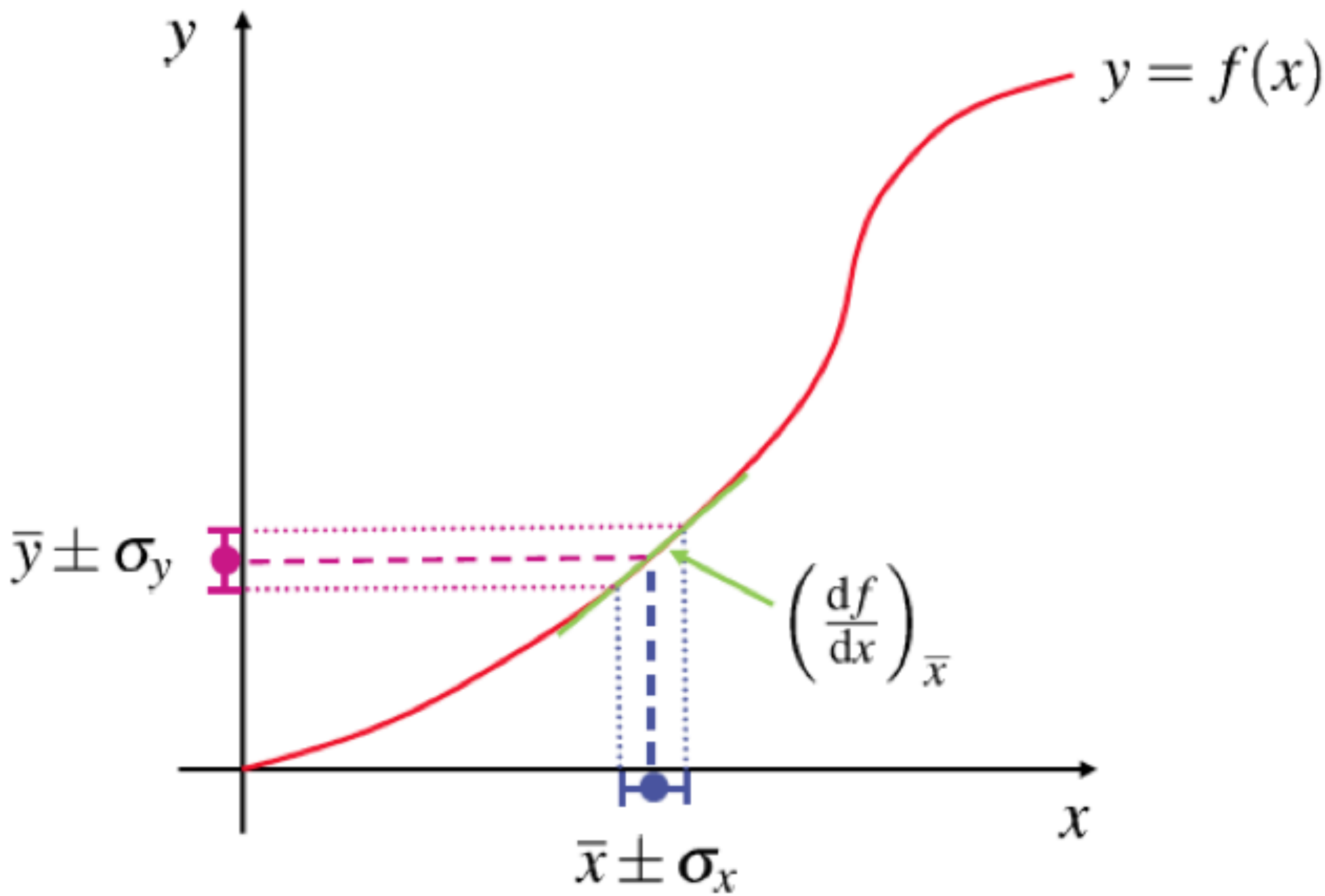
Correlation matrix of the p_T bins:



$$\rho_{i,j} = \frac{V_{i,j}}{\sigma_i \sigma_j}, \qquad V = \text{covariance matrix}$$

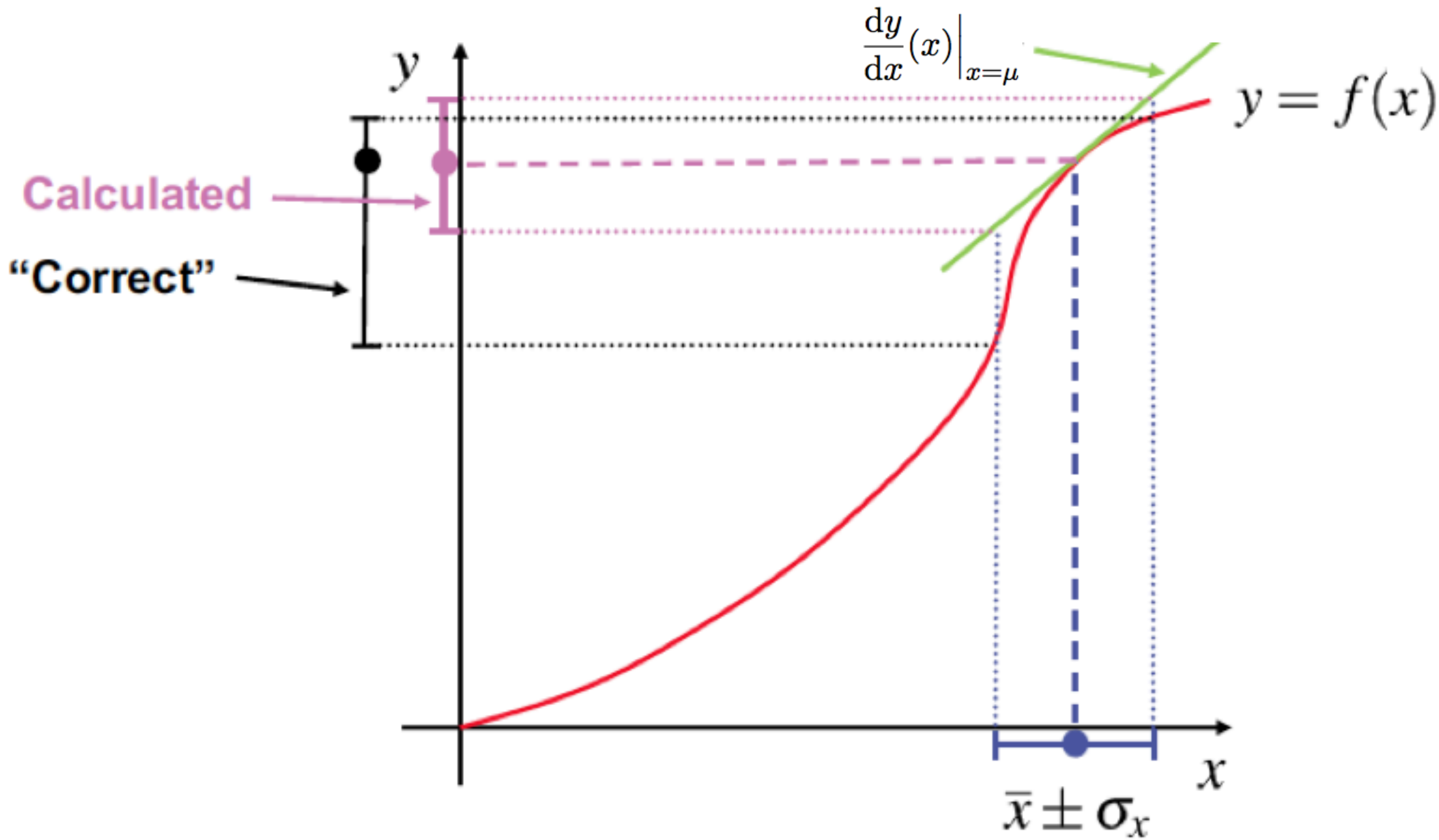
Error Propagation

Linear Error Propagation: Sometimes Applicable …



Function sufficiently linear within ±σ: linear error propagation applicable

Linear Error Propagation: Sometimes Not Applicable ...



In this situation linear error propagation is not applicable

Linear Error Propagation

Consider a measurement of values x_i and their covariances:

$$\vec{x} = (x_1, x_2, \ldots, x_n) \qquad V_{ij} = \mathrm{cov}[x_i, x_j]$$

Let y be a function of the x_i: $\quad y = f(\vec{x})$

What is the variance of y?

Approach: Taylor expansion of y around $\vec{\mu}$ where $\quad \mu_i = E[x_i]$

In practice we estimate μ_i by measured value x_i

$$V[y] \equiv \sigma_y^2 = E[y^2] - E[y]^2$$

Linear Error Propagation Formula

Taylor expansion: $$y(\vec{x}) \approx y(\vec{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)$$

$E[y]$ is easy: $$E[y] \approx y(\vec{\mu}) \qquad \text{as} \quad E[x_i - \mu_i] = 0$$

$E[y^2]$: $$E[y^2(\vec{x})] \approx y^2(\vec{\mu}) + 2y(\vec{\mu}) \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} E[x_i - \mu_i]$$
$$+ E\left[\left(\sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)\right)\left(\sum_{j=1}^{n} \left[\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} (x_j - \mu_j)\right)\right]$$
$$= y^2(\vec{\mu}) + \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} V_{ij}$$

Thus: $$\boxed{\sigma_y^2 = \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} V_{ij}}$$

Matrix Notation

Let vector A be given by $\vec{A} = \vec{\nabla} y$, i.e., $A_j = \left(\frac{\partial y}{\partial x_j} \right)_{\vec{x}=\vec{\mu}}$

Then: $$\sigma_y^2 = \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij} \quad = \quad A^T V A$$

Example: $$y = \frac{x_1}{x_2}, \quad A = \begin{pmatrix} 1/x_2 \\ -x_1/x_2^2 \end{pmatrix}$$

$$\sigma_y^2 = \left(\frac{1}{x_2}, -\frac{x_1}{x_2^2} \right) \begin{pmatrix} \sigma_1^2 & \mathrm{cov}[x_1, x_2] \\ \mathrm{cov}[x_1, x_2] & \sigma_2^2 \end{pmatrix} \begin{pmatrix} \frac{1}{x_2} \\ -\frac{x_1}{x_2^2} \end{pmatrix}$$

$$= \left(\frac{1}{x_2}, -\frac{x_1}{x_2^2} \right) \begin{pmatrix} \frac{\sigma_1^2}{x_2} - \frac{x_1}{x_2^2} \mathrm{cov}[x_1, x_2] \\ \frac{1}{x_2} \mathrm{cov}[x_1, x_2] - \frac{x_1}{x_2^2} \sigma_2^2 \end{pmatrix} = \frac{1}{x_2^2} \sigma_1^2 + \frac{x_1^2}{x_2^4} \sigma_2^2 - 2 \frac{x_1}{x_2^3} \mathrm{cov}[x_1, x_2]$$

$$\rightarrow \quad \frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} - 2 \frac{\mathrm{cov}[x_1, x_2]}{x_1 x_2} = \frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} - 2 \frac{\rho \sigma_1 \sigma_2}{x_1 x_2}$$

Linear Error Proportion: Examples

$$y = ax \quad \rightarrow \quad \sigma_y^2 = a^2 \sigma_x^2 \qquad \text{i.e. } \sigma_y = |a| \sigma_x$$

$$y = x^n \quad \rightarrow \quad \frac{\sigma_y^2}{y^2} = n^2 \frac{\sigma_x^2}{x^2} \qquad \text{i.e. } \frac{\sigma_y}{y} = |n| \frac{\sigma_x}{x}$$

$$y = x_1 + x_2 \quad \rightarrow \quad \sigma_y^2 = \sigma_1^2 + \sigma_2^2 + 2\text{cov}[x_1, x_2]$$

$$y = x_1 - x_2 \quad \rightarrow \quad \sigma_y^2 = \sigma_1^2 + \sigma_2^2 - 2\text{cov}[x_1, x_2]$$

$$y = x_1 x_2 \quad \rightarrow \quad \frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} + 2\frac{\text{cov}[x_1, x_2]}{x_1 x_2}$$

Sanity checks:

Average of fully correlated measurements:

$$y = \frac{1}{2}(x_1 + x_2), \; \sigma_1 = \sigma_2 \equiv \sigma, \; \rho = 1 \quad \rightsquigarrow \quad \sigma_y = \sigma$$

Difference of fully correlated measurements:

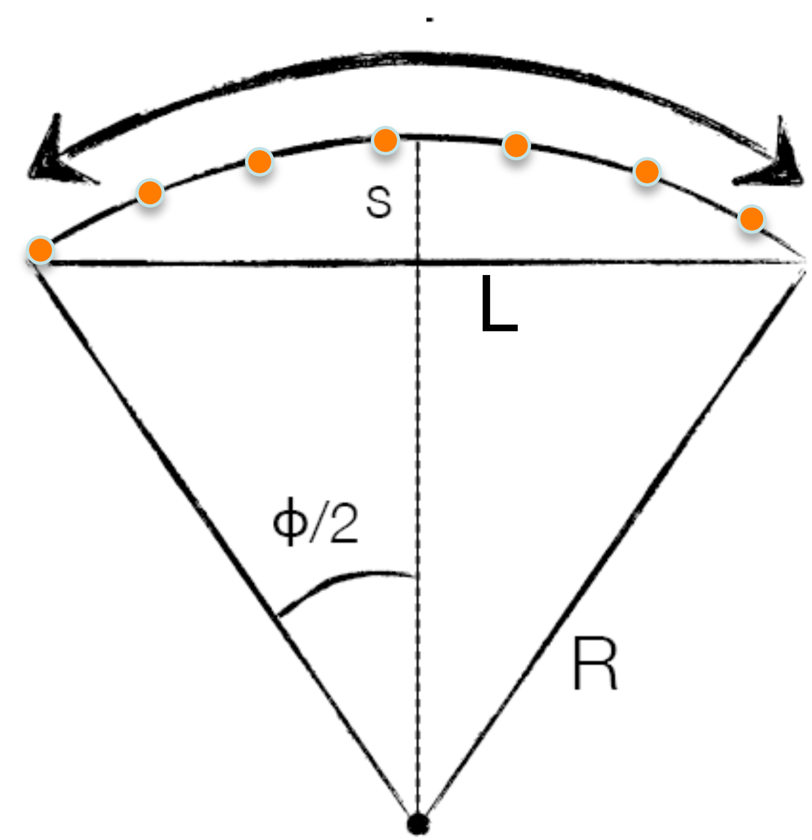
$$y = x_1 - x_2, \; \sigma_1 = \sigma_2 \equiv \sigma, \; \rho = 1$$

$$\rightsquigarrow \quad \sigma_y^2 = 2\sigma^2 - 2\sigma^2 = 0$$

Concrete Example: Momentum Resolution in Tracking

Charged particle moving in constant magnetic field:

$$p_T/\mathrm{GeV} = 0.3 \times B/\mathrm{Tesla} \times R/\mathrm{m}$$



Measurements of space points yields Gaussian uncertainty for sagitta *s* which is related to p_T as

$$R = \frac{L^2}{8s}, \quad p_T = 0.3B\frac{L^2}{8s}$$

Momentum resolution:

$$\frac{\sigma_{p_T}}{p_T} = \frac{\sigma_s}{s} = \frac{8p_T}{0.3BL^2}\sigma_s$$

Important features:

- Relative momentum uncertainty proportional to momentum
- Relative uncertainty prop. to uncertainty of coordinate measurement

Example:
ATLAS nominal resolution

$$\left(\frac{\sigma_{p_T}}{p_T}\right)^2 = \underbrace{0.001^2}_{\text{multiple scattering}} + \underbrace{(0.0005p_T)^2}_{\text{track uncertainty}}$$

Linear Error Propagation for Uncorrelated Measurements

Special case: the x_i are uncorrelated, i.e., $V_{ij} = \delta_{ij}\sigma_i^2$:

$$\sigma_y^2 = \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]^2_{\vec{x}=\vec{\mu}} \sigma_i^2$$

These formulas are exact only for linear functions.

Approximation breaks down if function is nonlinear over a region comparable in size to the σ_i.

Linear Error Propagation: Generalization from $\mathbb{R}^n \to \mathbb{R}$ to $\mathbb{R}^n \to \mathbb{R}^m$

Generalization: Consider set of *m* functions:

$$\vec{y}(\vec{x}) = (y_1(\vec{x}), y_2(\vec{x}), \ldots, y_m(\vec{x}))$$

Then:

$$\mathrm{cov}[y_k, y_l] \equiv U_{kl} \approx \sum_{i,j=1}^{n} \left[\frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}$$

In matrix notation:

$$U = AVA^T \qquad A_{ij} = \left[\frac{\partial y_i}{\partial x_j} \right]_{\vec{x}=\vec{\mu}}$$

Reduction of the Standard Deviation for Repeated Independent Measurements

Consider the average of n independent observation x_i:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Expectation values and variance of the measurements:

$$E[x_i] = \mu_i \qquad V[x_i] = \sigma^2$$

Standard deviation of the mean:

$$V[\bar{x}] = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 = \frac{1}{n} \sigma^2 \qquad \to \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Standard deviation of the mean decreases as $1/\sqrt{n}$

Example: Photon Energy Measurements

The energy resolution of a γ-ray detector used to investigate a decaying nuclear isotope is 50 keV.

- If only one photon is detected the energy of the decay is known to 50 keV
- 100 collected decays: energy of the decay known to 5 keV
- To reach 1 keV one needs to observe 2500 decays

Averaging Uncorrelated Measurements

Consider two uncorrelated measurements: $x_1 \pm \sigma_1, \; x_2 \pm \sigma_2$

Linear combination:

$$y = w_1 x_1 + w_2 x_2 \qquad \sigma_y^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2$$

Now choose the weights such that σ_y^2 is minimal (under the condition $w_1 + w_2 = 1$):

$$\frac{\partial}{\partial w_i} \sigma_y^2 = 0 \quad \rightarrow \quad w_i = \frac{1/\sigma_i^2}{1/\sigma_1^2 + 1/\sigma_2^2}$$

And for the uncertainty of y we obtain (linear error propagation):

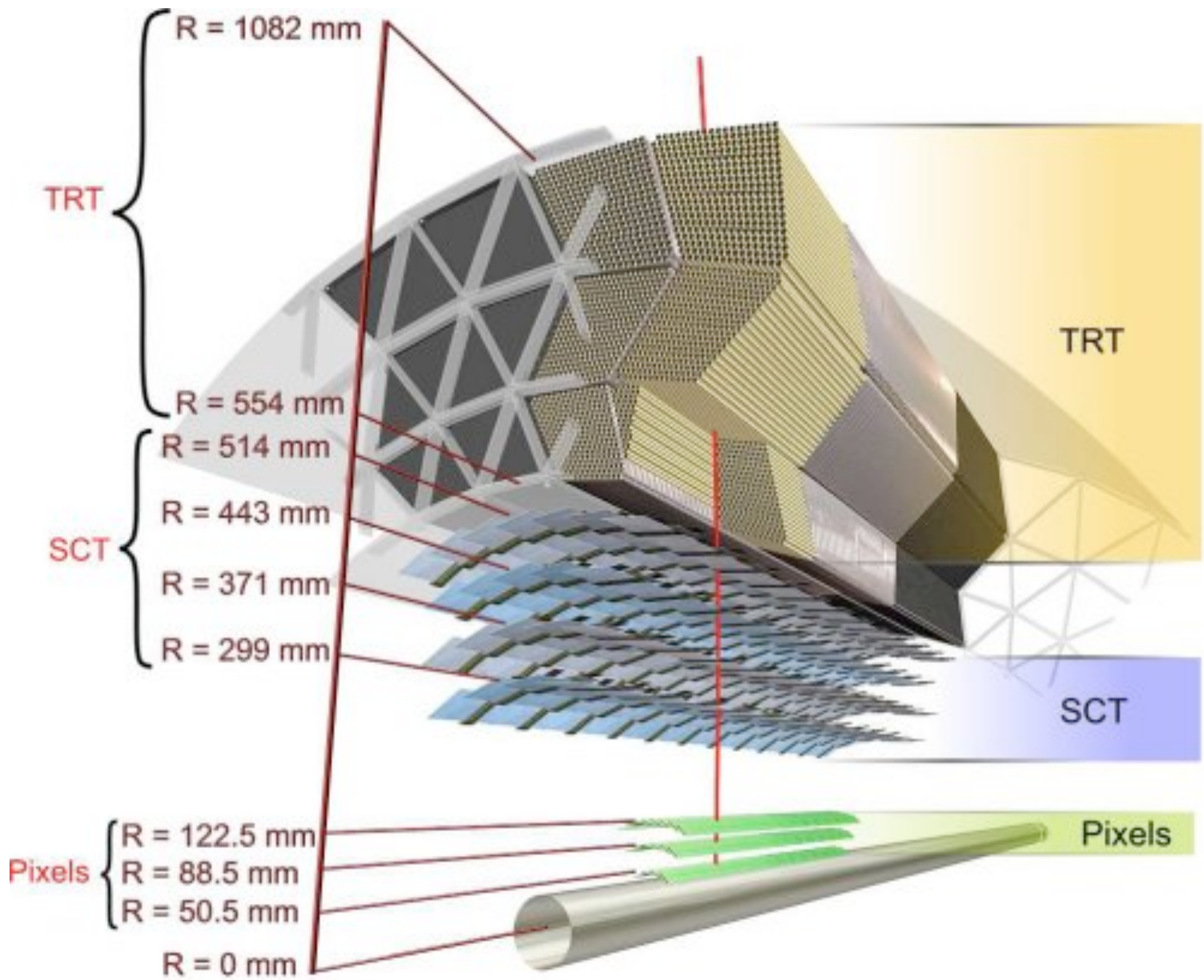
$$\frac{1}{\sigma_y^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

In general, for n uncorrelated measurements:

$$y = \sum_{i=1}^{n} w_i x_i, \qquad w_i = \frac{1/\sigma_i^2}{\sum_{j=1}^{n} 1/\sigma_j^2}, \qquad \frac{1}{\sigma_y^2} = \sum_{j=1}^{n} \frac{1}{\sigma_j^2}$$

Example: Averaging Uncorrelated Measurements

p_T of a particle in three subsystems of the ATLAS detector:



detector	p_T (GeV)
pixel detector	20 ± 2
semiconductor tracker	21 ± 1
transition radiation tracker	22 ± 4

Weighted average:

$$(20.86 \pm 0.87)\ \text{GeV}$$

$$p_T = \frac{\frac{20\ \text{GeV}}{4\ \text{GeV}^2} + \frac{21\ \text{GeV}}{1\ \text{GeV}^2} + \frac{22\ \text{GeV}}{16\ \text{GeV}^2}}{\frac{1}{4\ \text{GeV}^2} + \frac{1}{1\ \text{GeV}^2} + \frac{1}{16\ \text{GeV}^2}} = 20.86\ \text{GeV}$$

$$\sigma_{p_T} = \left[\frac{1}{4\ \text{GeV}^2} + \frac{1}{1\ \text{GeV}^2} + \frac{1}{16\ \text{GeV}^2}\right]^{-1/2} = 0.87\ \text{GeV}$$

Weighted Average from Bayesian Approach

Consider two measurements μ_1 and μ_2 with Gaussian uncertainties σ_1 and σ_2. In a Bayesian approach the probability distribution for the true value x is given by

$$p(x) \propto L(\mu_1, \mu_2 | x) \pi(x)$$

Assuming a flat prior $\pi(x) \equiv 1$ and independence of the two measurements one obtains

$$\begin{aligned} p(x) &\propto L(\mu_1|x) L(\mu_2|x) \\ &= G(\mu_1; x, \sigma_1) G(\mu_2; x, \sigma_2) \\ &\propto \exp\left[-\frac{1}{2}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(x-\mu_2)^2}{\sigma_2^2}\right)\right] \end{aligned}$$

The product of the two Gaussians gives a Gaussian with mean

$$\mu = w_1 \mu_1 + w_2 \mu_2 \quad \text{where } w_i = \frac{1/\sigma_i^2}{1/\sigma_1^2 + 1/\sigma_2^2}$$

and standard deviation

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

→ same result as before

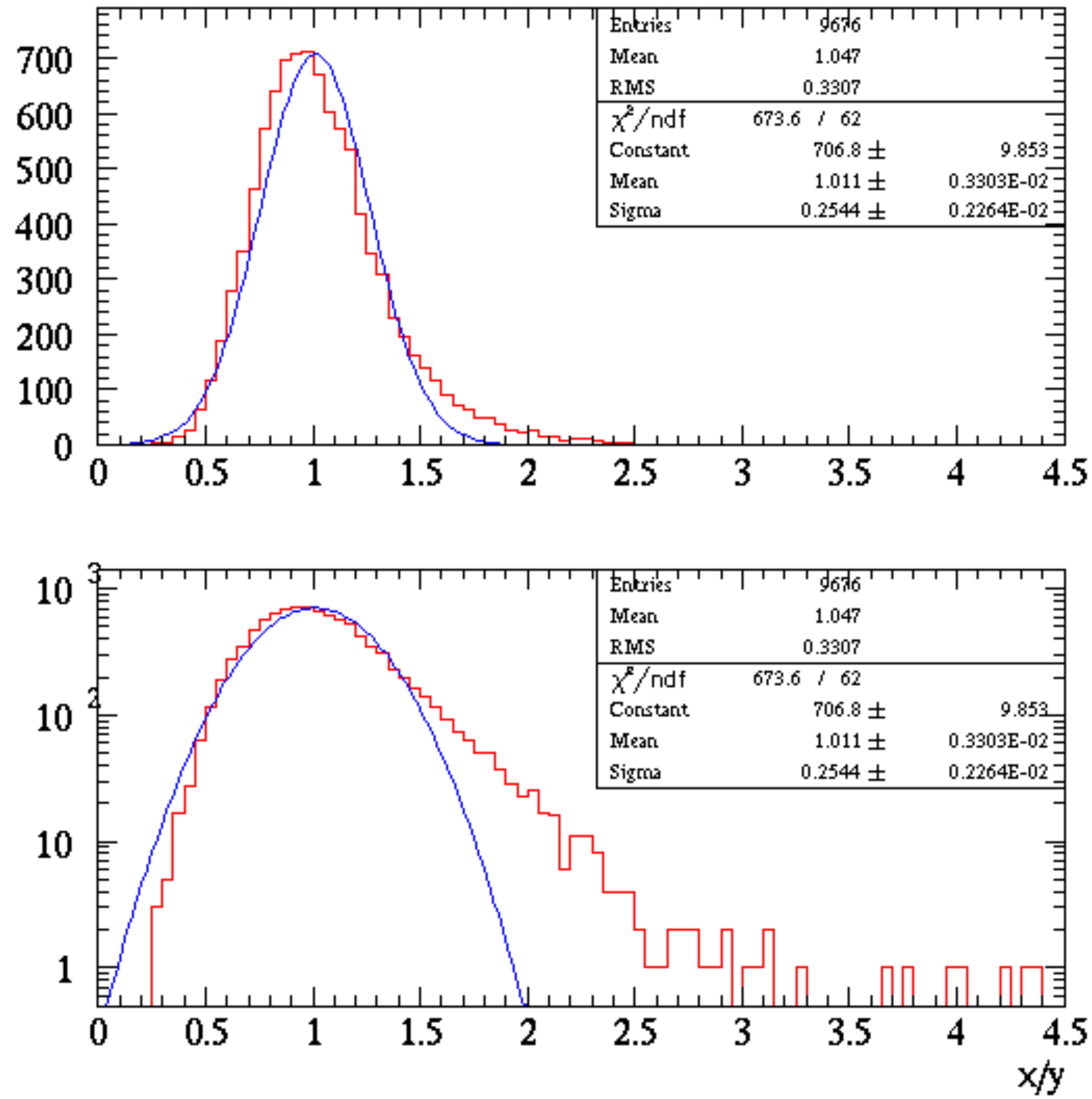
Monte Carlo Error Propagation

Example:
Ratio of two Gaussian distributed quantities

$$x = 5 \pm 1$$

$$y = 5 \pm 1$$

Approach: draw values for *x* and *y* many times and fill histogram with ratios



Standard linear error prop.:

$$R = 1 \pm 0.28$$

Mean and rms of histogram:

$$R = 1.05 \pm 0.33$$

Rule of thumb: ratio of two Gaussians will be approximately Gaussian if fractional uncertainty is dominated by numerator, and denominator cannot be small compared to numerator